

The Genome Organization of the Nidovirales: Similarities and Differences between Arteri-, Toro-, and Coronaviruses

Antoine A. F. de Vries, Marian C. Horzinek,
Peter J. M. Rottier, and Raoul J. de Groot¹

Virology Unit, Department of Infectious Diseases and Immunology, Veterinary Faculty,
Utrecht University, Yalelaan 1, 3584 CL Utrecht, The Netherlands

Viruses in the families Arteriviridae and Coronaviridae have enveloped virions which contain nonsegmented, positive-stranded RNA, but the constituent genera differ markedly in genetic complexity and virion structure. Nevertheless, there are striking resemblances among the viruses in the organization and expression of their genomes, and sequence conservation among the polymerase polyproteins strongly suggests that they have a common ancestry. On this basis, the International Committee on Taxonomy of Viruses recently established a new order, Nidovirales, to contain the two families. Here, the common traits and distinguishing features of the Nidovirales are reviewed. © 1997 Academic Press

KEY WORDS: arterivirus; coronavirus; torovirus; polyprotein processing; RNA recombination.

INTRODUCTION

The Nidovirales (summarized in Table 1) is a newly established order comprising the families Arteriviridae (genus *Arterivirus*) and Coronaviridae (genera *Coronavirus* and *Torovirus*). Species in the genus *Coronavirus* can be grouped into three clusters on the basis of serological and genetic properties (1). Two torovirus species have been recognized: the equine and bovine toroviruses (ETV, Berne virus; and BoTV, Breda virus). In addition, a human torovirus is thought to exist (2) and we have recently identified a porcine torovirus (PoTV) (Kroneman *et al.*, unpublished). The genus *Arterivirus* presently contains four species.

Despite considerable differences in genetic complexity and virion architecture, coronaviruses, toroviruses, and arteriviruses are strikingly similar in genome organization and replication strategy (3) (Fig. 1). The name Nidovirales (from the Latin *nidus*, nest) refers to the 3' coterminal nested set of subgenomic (sg) viral mRNAs that is produced during infection. Sequence similarities, although mostly restricted to the lb poly-

protein (POL1b) from which the replicase-associated proteins are derived, suggest that the Nidovirales have evolved from a common ancestor. Apparently their divergence has been accompanied by extensive genome rearrangements through heterologous RNA recombination.

Here, we review the common traits and distinguishing features of the genome organization, gene expression, and evolution of the Nidovirales. Other reviews are references 3 to 9 and the different models proposed for sg mRNA synthesis are discussed in references 8 to 10.

VIRION ARCHITECTURE AND STRUCTURAL PROTEINS

The phylogenetic relationship among arteriviruses, toroviruses, and coronaviruses is not apparent from their morphology. Coronavirions are roughly spherical, 100–120 nm in diameter, with a fringe of c. 20-nm-long petal-shaped spikes. Some group II coronaviruses exhibit a second fringe of smaller surface projections about 5 nm in length. Torovirus particles are pleiomorphic, measuring 120 to 140 nm in their

¹To whom correspondence and reprint requests should be addressed. Fax: +31-30-2536723. E-mail: R.Groot@vetmic.dgk.ruu.nl.

TABLE 1

Order:	Nidovirales					
Family:	Arteriviridae			Coronaviridae		
Genus:	<i>Arterivirus</i>			<i>Torovirus</i>		<i>Coronavirus</i>
Species:	Equine arteritis virus	EAV	Equine torovirus	ETV	Transmissible gastroenteritis virus	TGEV
	Porcine reproductive and respiratory syndrome virus	PRRSV	Bovine torovirus	BoTV	Feline coronavirus	FCoV
	Lactate dehydrogenase-elevating virus	LDV	Porcine torovirus	PoTV	Canine coronavirus	CCV
	Simian hemorrhagic fever virus	SHFV			Human coronavirus	HCV 229E
					Porcine epidemic diarrhea virus	PEDV
					Mouse hepatitis virus	MHV
					Bovine coronavirus	BCV
					Human coronavirus	HCV OC43
					Porcine hemagglutinating encephalomyelitis virus	HEV
					Sialoacryoadentitis virus	SADV
					Turkey coronavirus	TCV
					Infectious bronchitis virus	IBV

largest axis; spherical, oval, elongated, and kidney-shaped virions have been described. The surface projections on torovirus virions closely resemble coronavirus peplomers (11). Arterivirions are only 50–70 nm in diameter and lack large surface projections. Instead, cup-like structures with a diameter of 10 to 15 nm have been observed (12). The difference in virion architecture become even more apparent when comparing the nucleocapsid structures. That of coronaviruses is a loosely wound helix (13), that of toroviruses is a compact tubular structure (11), and that of arteriviruses is isometric, about 25–35 nm in diameter, and possibly icosahedral (12). The nucleocapsid proteins (N) differ considerably in size (c. 50, 19, and 14 kDa for corona-, toro-, and arteriviruses, respectively) and amino acid sequence.

The compositions of the viral envelopes also differ. Coronavirus membranes contain: (i) 180- to 220-kDa spike protein (S), (ii) 25- 30-kDa triple-spanning membrane protein M, and (iii) c. 10-kDa transmembrane protein E, a minor virion component but essential for virus assembly (14,15). The small surface projections of group II coronaviruses are dimers of a 65-kDa class I membrane protein, the hemagglutinin-esterase (HE), possibly acquired by heterologous RNA recombination (16,17).

Toroviruses also specify M and S proteins of 26 and 180 kDa, respectively. Although different in sequence, the M and S proteins of toro- and coronaviruses are alike in size, structure, and function. The M proteins have a similar triple-spanning membrane topology

(18), and the heptad repeats, indicative of a coiled-coil structure in the spike proteins of coronaviruses (19), are also present in the torovirus peplomer (20). Thus, the S and M genes of these viruses may well be phylogenetically related (6,18,20). Puzzlingly, toroviruses seem to lack a homologue for the E protein, which could indicate a difference in assembly. We have found recently that BoTV virions contain a third membrane protein, the 65-kDa hemagglutinin-esterase (145).

The structural proteins of arteriviruses are unrelated to those of the Coronaviridae. There is a basic set of three envelope proteins (21-24). (i) a 16- to 20-kDa nonglycosylated membrane protein (M) which traverses the membrane three times and thus structurally resembles the M protein of corona- and toroviruses, (ii) a heterogeneously N-glycosylated triple-spanning protein (designated G_L for EAV) of variable size, and (iii) a class I glycoprotein of 25–30 kDa (designated G_S for EAV) which is a minor virion component. The G_L and M proteins associate into disulphide-linked heterodimers and probably form the cup-like structures on the virion surface (24–26).

GENES AND REGULATORY ELEMENTS

Overall Genome Structure

Nidoviral genome RNA is single-stranded, infectious, polyadenylated (27–29), and, at least for arteri- and coronaviruses, 5' capped (30,31). Nucleotide sequences

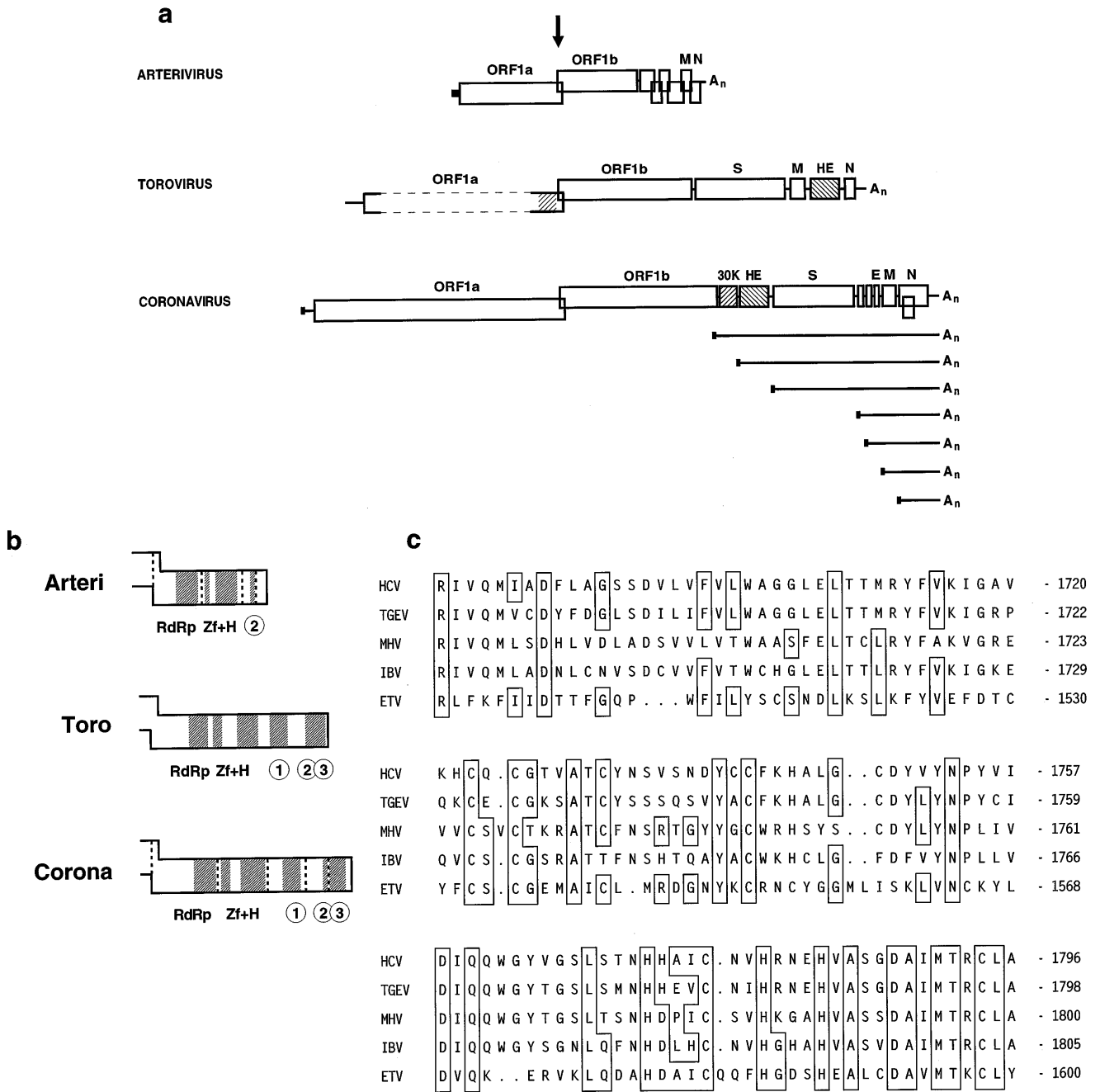


FIG. 1. (a) Scale representation of archetypical Nidovirales genomes. The torovirus genome organization is based on combined data for ETV and BoTV, and that of coronavirus is typical for a group I member (Table 1). The 5' ends of ORF1b have been aligned. The bottom panel illustrates the 3' coterminal nested set of mRNAs produced during coronavirus infection. ORFs are represented by boxes. The coding assignments are also indicated. Hatched boxes represent the ORFs for HE and the 30-kDa ns2a protein of group I coronaviruses and related sequences in the torovirus genome. The arrow indicates the position of the pseudoknot structures required for translational read-through of ORF1b. The 5' leader sequences are depicted by a small black box. Poly(A) tails are indicated by A_n. (b) Sequence conservation in the POL1b polyproteins. Conserved domains are indicated by hatching. RdRp, Zf, and H indicate the RNA-dependent RNA polymerase, zinc finger, and helicase motifs, respectively. Domains 1-3 indicate conserved regions for which as yet no function has been suggested. Motif 2 corresponds to the previously described CVL domain. Motif 1 has not been described before. Bracketed lines indicate (predicted) proteolytic cleavage sites (for details see text). (c) Sequence conservation in motif 1. Sequences were taken from (32,34,36,41,144). Residues conserved between corona- and toroviruses are boxed.

are known for the complete RNA of coronaviruses MHV, IBV, TGEV, and HCV 229E and arteriviruses EAV, LDV, and PRRSV (32–39) and for parts of RNA of several other Nidovirales, including ETV strain Berne (40,41) and SHFV (Godeny *et al.*, in press). The size of the arterivirus genome is from 13 to 15 kb. The genomes of toroviruses and coronaviruses are considerably larger (up to 31 kb) and include the largest known RNA genomes. Despite the differences in genetic complexity and gene composition, the genome organizations of arteri-, toro-, and coronaviruses are remarkably similar. More than two-thirds of each genome are taken up by two huge overlapping open reading frames (ORFs), designated ORF1a and 1b. The more downstream, ORF1b, is only expressed after translational read-through via a -1 frameshift mediated by a pseudoknot structure (42). The polypeptides encoded by these ORFs are proteolytically cleaved by virus-encoded proteinases to yield the proteins involved in viral RNA synthesis.

Downstream of ORF1b, there are four to nine genes that encode the structural proteins and, at least for coronaviruses, a number of nonstructural proteins. These genes are expressed from a 3' coterminal nested set of sg mRNAs (8,40,43,44). Although these mRNAs are structurally polycistronic, translation is restricted to the unique 5' sequences not present in the next smaller RNA of the set. Cells infected by arteriviruses or coronaviruses contain negative-stranded RNAs which correspond to each mRNA and which may serve as templates for transcription (45–49).

Sequence Elements Regulating Transcription

Each transcription unit (comprising one or more genes expressed from a single mRNA species) is preceded by a short consensus sequence, the complement of which is thought to function as a promoter: the transcription-associated sequence (TAS) (3,10,50). The relative strength of coronavirus promoters is influenced by the primary structure of the TAS (10,50,51) and the presence downstream of other TASs. In general, downstream TASs have a negative effect on transcription levels from upstream sites (52–54). For MHV, host proteins of 35 and 38 kDa have been identified that specifically bind to the TAS and may serve as transcription factors (9,55,56).

The sg mRNAs of corona- and arteriviruses carry a 5' leader sequence of 55–92 and about 200 nt, respectively, which are derived from the 5' ends of the viral genomes. The mRNA synthesis thus requires, at least

at one point, a discontinuous transcription event (43,44). The fusion of “leader” and “body” sequences occurs within or in close proximity to the TAS (10,49,57,58). Puzzlingly, the torovirus mRNAs seem to lack an extensive 5' leader sequence (40,59). Thus if the use of a leader sequence evolved before the divergence of the Nidovirales, toroviruses must have lost their leader relatively recently. The close evolutionary relationship between toro- and coronaviruses suggests that this event took place after the Coronaviridae and Arteriviridae diverged. Alternatively, the common ancestor of the Nidovirales may have used a leader-independent transcription mechanism and arteri- and coronaviruses acquired a 5' leader independently. In either view, the addition of noncontiguous leader sequences would not be a mechanistically important aspect of mRNA synthesis (as suggested by the “leader-primed” transcription model) (8) but rather a modification of a common transcription scheme, based primarily on transcriptase-promoter recognition (9,60). What then is the function of the leader sequence? Perhaps the discontinuous transcription seen in arteri- and coronaviruses has evolved merely to provide each viral mRNA with a translational enhancer, allowing efficient competition with host mRNAs for the cellular translational machinery. Indeed, there is evidence that the coronavirus leader sequence stimulates viral translation *in cis*, possibly in conjunction with a virus-specified or virus-induced factor (61).

For a complete understanding of Nidovirales transcription-initiation, studies on torovirus mRNA synthesis will be pivotal. In fact, the existence of a small torovirus leader RNA cannot entirely be excluded. Sequence analysis of ETV defective interfering RNAs, combined with results of primer extension studies, suggest that a TAS is present at the extreme 5' end of the viral genome which could give rise to a leader of approximately 8 nt (59).

5' and 3' Nontranslated Regions

The promoters required for genome replication are commonly found at the 5' and 3' ends of the genome. Coronaviruses have nontranslated regions (NTRs) ranging from 0.2 to 0.5 kb (5') and from 0.3 to 0.5 kb (3'). Their primary structure is poorly conserved among the different subgroups. Deletion mapping studies using synthetic DI RNAs suggest that for the group II coronaviruses, about 0.5 kb of each end of the genome is required for replication, implying that promoter elements may extend into ORF1a and the N gene

(10,62–66). All coronavirus genome RNAs have the sequence 5' U/GGGAAGAGC 3' about 70 nt upstream of the poly(A) tail (67,68). The strict conservation of this sequence element suggests that it has a role in replication. Surprisingly, however, the 3' most 55 nt of the 3' NTR of MHV appear to be sufficient to drive minus-strand synthesis (69).

The 3' NTRs of toroviruses are about 0.3 kb. The 5' NTR of ETV strain Berne is 0.8 kb (59) but the lengths of 5' NTRs of other toroviruses are unknown. The 5' NTRs of arteriviruses are about 0.2 kb and, unlike those of coronaviruses, consist almost entirely of the leader (37–39,70). The 3' NTRs of arteriviruses are also short, ranging from 59 to 151 nt, and conserved sequence elements have not been found.

POLYPROTEIN PROCESSING: THE POLYMERASE GENE

The overlapping ORFs 1a and b found at the 5' end of the nidoviral genome are frequently referred to as the "polymerase gene." However, there is little doubt that the processing of the encoded polyproteins yields proteins required for RNA synthesis as well as a number of products involved in other aspects of virus replication. The 1a and 1b polyproteins of coronaviruses are 3951 to 4492 and 2682 to 2714 residues long, respectively. POL1b of ETV strain Berne consists of 2289 residues; only limited sequence data are available for torovirus ORF1a. The polyproteins of arteriviruses are much smaller, with lengths of 1727–2396 (POL1a) and 1411–1459 (POL1b) residues.

Amino acid sequence comparisons show that the 1b polyproteins of corona-, toro-, and arteriviruses are basically colinear (37,41) (Fig. 1b). The sequence conservation between the more closely related corona- and toroviruses is clustered in six domains, four of which are also found in the arterivirus POL1b: the "classical" RNA-dependent RNA polymerase (RdRp) and helicase (H) domains, which are also present in the polymerases of most other viruses, a zinc finger motif (zf), and a short region of 80–100 residues, which has not yet been identified in other viral polymerases and was called the "coronavirus-like" (CVL) domain (3) (motif 2 in Fig. 1b).

Processing of Coronavirus POL1A Polyproteins by Papain-like Proteinases

There is little sequence conservation among the N-termini of the POL1a polyproteins of the three

coronavirus subgroups. Size differences can mostly be attributed to these regions (Fig. 2) and sequence similarities are limited to papain-like cysteine proteinase (pcp) domains (33,34,36,71). POL1a of HCV 229E, TGEV, FIPV (subgroup I), and MHV (subgroup II) have two pcp domains, whereas that of IBV (subgroup III) contains a single pcp domain. These pcp seem to be involved in the processing of the N-termini of the 1a polyproteins.

The proteolytic cleavage of the N-terminus of the coronavirus 1a polyprotein has been studied in most detail for MHV. *In vitro* translation of genomic RNA gave products of 28 and 220 kDa and the production of p28 was sensitive to proteinase inhibitors, suggesting that it arose by a proteolytic cleavage (72). p28 was also detected in MHV-infected cells (73). Partial peptide mapping revealed that p28 is derived from the N-terminus of POL1a (74). Baker *et al.* (75) subsequently showed that the proteolytic activity responsible for the production of p28 mapped to residues 1223–1695 of POL1a which contains the N-terminal-most pcp domain (pcp1) (33). Mutagenesis showed that any change of either Cys¹¹³⁷ or His¹²⁸⁸ (Cys¹¹²¹ and His¹²⁷² of MHV-A59) (35,76) resulted in the loss of proteinase activity, suggesting that these residues form the catalytic dyad (77). Cleavage to give p28 was at an RGV motif at the G²⁴⁷/V²⁴⁸ dipeptide bond (78,79), and presumably occurred *in cis* (75). Reactions of specific antisera raised against different regions of MHV POL1a with potential cleavage products with apparent molecular weights of 65, 50, 240, and 290 kDa in MHV-infected cells (80,81) showed that processing of the N-terminus of POL1a involves multiple cleavage events. p65 is thought to be immediately adjacent to p28 (81,82). Gao *et al.* (82) reported that p65 of MHV strain JHM is generated from a p72 precursor, but this precursor has not been observed by others studying MHV strain A59 (81). Kinetic analysis suggests that p290 is a precursor to p50 and p240. A provisional map of the POL1a region of MHV is shown in Fig. 2. The proteinases involved in the release of p65, p50, and p240 have not yet been identified. Although some authors have implicated pcp1 in the cleavage of p65 (76) this is disputed by others (82).

Only limited data are available on the processing of the N-terminus of POL1a of IBV. Using monospecific antisera raised against residues 49–514 or 247–599, Liu *et al.* (83) detected a 87-kDa product in IBV-infected cells. It is not known if IBV p87 represents the N-terminal cleavage product or if an additional smaller product is released from the N-terminus of POL1a. p87 was also found upon *in vivo* expression of the N-terminal

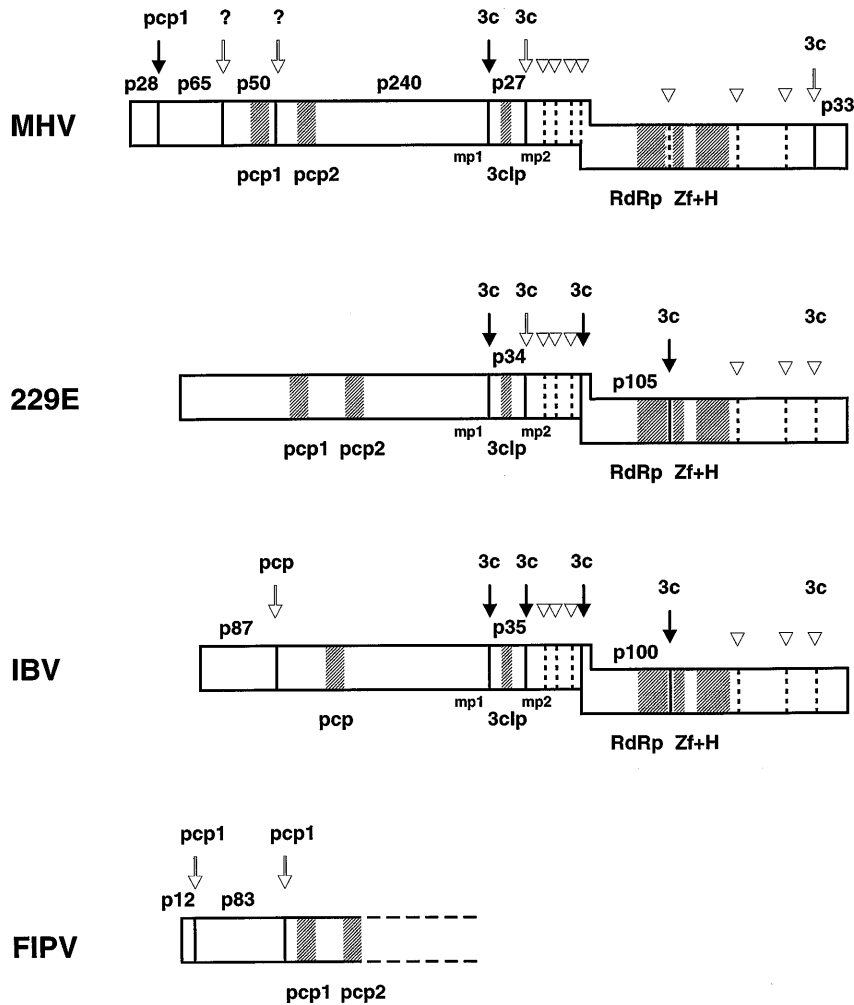


FIG. 2. Proteolytic processing of the coronavirus polyproteins POL1a and POL1ab. Provisional cleavage maps were constructed on the basis of the combined data discussed in the text. POL1a and POL1b sequences are indicated by boxes. Papain-like (pcp) and 3C-like cysteine proteinase (3clp) domains are indicated by shading, as are the RNA dependent RNA polymerase (RdRp), zinc finger (Zf), and helicase domains (H). Also shown are the hydrophobic domains, mp1 and mp2, that flank 3clp. Cleavage sites that have been identified experimentally either by protein sequence analysis or by site-directed mutagenesis are indicated by black arrows. White arrows indicate cleavages for which the exact cleavage site has not been determined. Cleavage products are designated after their apparent molecular weight as determined by SDS-PAGE. Proteinases involved in each cleavage event are given. Question marks indicate cleavages for which the proteinase has not yet been identified. Open arrowheads indicate predicted cleavage sites for 3clp.

1742 residues of IBV POL1a (83), which include the pcp domain (33,71). Interestingly, p87 was not detected after *in vivo* expression of a shorter N-terminal polypeptide of 1444 residues that lacked pcp, strongly suggesting that pcp is involved in the release of this product. Because p87 did not appear when the 1742-residue polypeptide was produced by *in vitro* translation, cellular factors may also be involved in this cleavage event. However, *in vivo* processing of this polypeptide was also inefficient, possibly because the pcp is located at the C-terminus of the 1742-residue expression prod-

uct and sequences downstream of this domain are required for optimal proteolytic activity.

In our laboratory, a monospecific antiserum, raised against the N-terminal 198 residues of the 1a polypeptide of FIPV, specifically recognized products of 12, 83, and 100 kDa in FIPV-infected cells. These products were also found upon *in vivo* expression of the N-terminal 1446 residues of FIPV POL1a containing the pcp1 domain. Kinetic analysis suggested that p12 and p83 are mature products with p100 as their precursor. p12 reacted with antiserum raised against the N-terminal

15 residues of POL1a, showing it to be the N-terminal-most cleavage product. pcp1 appears to be involved in the release of both p12 and p83. Substitution of the presumptive catalytic cysteine residue of pcp1 (Cys¹¹¹⁷), completely abolished proteolytic activity (Fig. 2; De Groot *et al.*, in preparation).

Processing of the Coronavirus Polyproteins 1a and 1b by the 3C-like proteinase

In contrast to the N-termini, the C-terminal third of coronavirus POL1a polyproteins are well conserved. All contain a proteinase domain flanked by hydrophobic regions, designated mp1 and mp2 (Fig. 2). This proteinase is related to the chymotrypsin-like serine proteases, but with a cysteine rather than a serine residue as the active site nucleophile (33,34,36,71,84). A similar situation exists in the 3C proteinases of picornaviruses and 3C-like proteinases of plant viruses (85).

The 3C-like proteinases (3clp) of coronaviruses are involved in the processing of the C-terminus of POL1a and of POL1ab. The results obtained for IBV, MHV, and HCV 229E differ only in details. The 3clp mediates at least four cleavage events. It autocatalytically excises itself from the polyprotein precursor, yielding products of 35, 27, and 34 kDa for IBV, MHV, and HCV 229E, respectively (86–89) (Fig. 2). The release of IBV 3clp (but not that of MHV) from a synthetic precursor *in vitro* was dependent on the presence of microsomal membranes and apparently required membrane-association of the flanking lipophilic domains (86,87). Lu *et al.* (88) proposed that because production of the MHV p27 *in vitro* was sensitive to dilution, the autocatalytic release of 3clp occurs mainly *in trans*. Protein sequence analysis identified Q³³³³/S³³³⁴ and Q²⁹⁶⁵/A²⁹⁶⁶ as the respective N-terminal cleavage sites of MHV p27 and HCV p34 with the Gln residues in the P1 position (87,89). p35 of IBV is generated by cleavage of QS dipeptides at positions 2779–2780 and 3086–3087 (86). The cleavage sites flanking 3clp are well conserved among the different coronaviruses.

Processing of the POL1ab polyprotein by 3clp also resulted in the production of a polypeptide of c. 100 kDa, containing the RdRp domain (90–92). The cleavage sites for IBV and HCV 229E were at the positionally conserved dipeptides Q³⁹²⁸/S³⁹²⁹ and Q⁴⁸⁶⁸/S⁴⁸⁶⁹ or Q⁴⁰⁶⁸/S⁴⁰⁶⁹ and Q⁴⁹⁹⁵/A⁴⁹⁹⁶, respectively, the N-terminal most of which are located in POL1a (Fig. 2). Processing leading to the release of the RdRp can occur *in trans*, both *in vitro* and *in vivo* (91,92).

Gorbalenya *et al.* (71) predicted that the catalytic site

of the IBV 3clp consists of a triad formed by His²⁸²⁰, Glu²⁸⁴³, and Cys²⁹²². The Cys and His residues are conserved in the 3clp domain of the other coronaviruses and their involvement in proteolysis has been confirmed by site-directed mutagenesis (86,87,89,91). Glu²⁸⁴³ is not part of the catalytic site. This residue is not conserved in other 3clp and substitution by Asn, Asp, or Gln did not affect proteolytic activity (91). In agreement with the assumed evolutionary relationship with cellular trypsin-like serine proteases, the coronavirus 3clp are sensitive to both serine and cysteine protease inhibitors (86,88). Moreover, substitution of the active site Cys by Ser yielded an IBV 3clp which was still partially active (86).

The cleavage sites of the coronavirus 3clp conform to the consensus XQZ, with X being a hydrophobic residue (L, V, I, M or F) and Z a small uncharged residue (S, A, G or C). These data provide experimental support to earlier predictions (33,71). Alignment of POL1ab sequences suggests that 3clp may cleave at seven additional conserved sites (Fig. 2). Cleavage at the sites in MHV POL1a would produce four extra polypeptides with predicted molecular weights of 33, 10, 34, and 15 kDa. The 33-kDa product would contain the hydrophobic domain mp2, whereas the 15-kDa product would be a cysteine-rich polypeptide resembling murine epidermal growth factor in sequence (71). Processing of POL1b would yield the RdRp and four other products. The zinc finger and helicase motifs would be in a product of about 67 kDa and the conserved motif 1 would be in a polypeptide of 59 kDa, whereas motifs 2 (the CVL domain) and 3 would be in products of 42 and 33 kDa, respectively (Figs. 1 and 2). The latter may correspond to a 33-kDa protein in lysates of MHV-infected cells which reacted with antiserum against the 14 C-terminal amino acids of POL1b (93).

Processing of the Arterivirus Polymerase Polyproteins

Most of what is known about arterivirus polyprotein processing stems from the work of Snijder and colleagues on EAV; only limited information is available for PRRSV and LDV. As for coronaviruses, most sequence variation occurs in POL1a. Processing of the N-terminus of POL1a is mediated by papain-like cysteine proteinases, whereas the C-terminus of POL1a and the conserved 1b polyprotein is probably processed by a 3C-like proteinase which is located at the

C-terminus of POL1a and flanked by hydrophobic domains (Fig. 3).

For both PRRSV and LDV (38,39), the N-terminus of POL1a contains two papain-like proteinase domains, pcp α and pcp β , which mediate their own release by cleavage *in cis* at C-terminal cleavage sites, giving rise to products nsP1 α and nsP1 β (Fig. 3) (94). The PRRSV and LDV leader proteinases share 48% sequence identity. For PRRSV, Cys⁷⁶ and His¹⁴⁶ are crucial for pcp α activity (94), whereas cleavage by pcp β was dependent on Cys²⁷⁶ and His³⁴⁵. For LDV, Cys⁷⁶ and Cys²⁶⁹ were identified as active site cysteines. The cleavage sites in POL1a have not been mapped but from the sizes of nsP1 α and nsP1 β , and from the results of deletion analyses, are predicted to be around position 170 for pcp α and between Tyr³⁸⁴ and Gly³⁸⁵ for PRRSV pcp β , and between Tyr³⁸⁰ and Gly³⁸¹ for LDV pcp β .

EAV is thought to have a single leader proteinase (37), corresponding to pcp β of LDV and PRRSV.

However, relicts of nsP1 α are still present in the N-terminus of EAV POL1a (94). The EAV pcp β homologue releases a 29-kDa protein, nsP1⁹⁵ (Fig. 3), apparently exclusively by cleavage *in cis* at G²⁶⁰/G²⁶¹. The results of site-directed mutagenesis suggested that Cys¹⁶⁴ and His²³⁰ form the catalytic dyad (95).

Four additional mature cleavage products were identified in lysates of EAV-infected cells (96) and were designated nsP2 to 5 (Fig. 3). The 61-kDa nsP2 protein is released by cleavage between Gly⁸³¹ and Gly⁸³² and the catalytic activity responsible is within the N-terminal 165 residues of nsP2 as this domain can induce cleavage at the 2/3 site *in trans* (97). Sequence comparisons suggested that the catalytic residues in the cysteine proteinase domain were Cys²⁷⁰ and His³³². Substitutions of these residues completely abolished proteolytic activity, but so did replacement of three other conserved cysteine residues (positions 319, 349, and 354). The N- and C-terminal sequences of nsP2 are highly

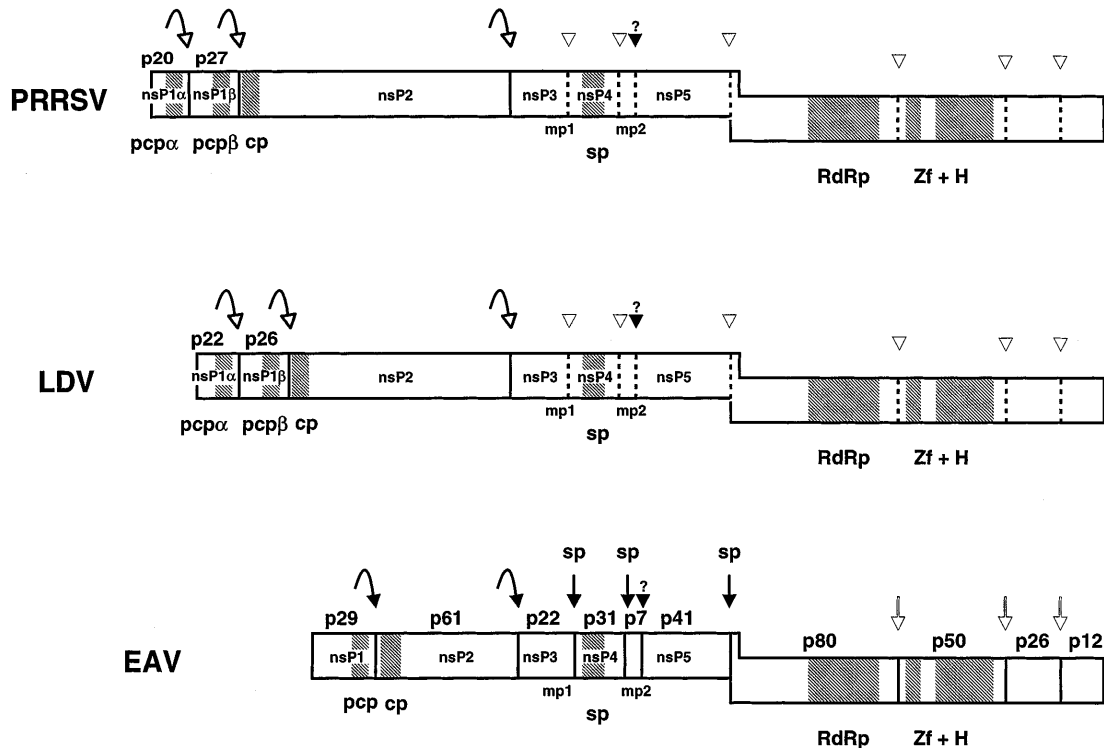


FIG. 3. Proteolytic processing of the arterivirus polyproteins POL1a and POL1ab. The (provisional) cleavage maps were constructed on the basis of the combined data discussed in the text. POL1a and POL1b sequences are indicated by boxes. POL1a cleavage products are numbered according to Snijder *et al.* (96). Also shown are the apparent molecular weights of the cleavage products. The papain-like proteinase domains (pcp) and the nsP2 cysteine (cp) and the nsP4 serine proteinases (sp) are indicated by shading as are the RNA-dependent RNA polymerase (RdRp), zinc finger (Zf), and helicase domains (H). Also shown are the hydrophobic domains, mp1 and mp2, that flank nsP4. Cleavage sites that have been identified experimentally are indicated by black arrows. White arrows indicate cleavages for which the exact cleavage site has not yet been determined. Cleavages performed by the serine proteinase are given. Arched arrows depict cleavages performed by the leader proteinases. Open arrowheads indicate predicted sp cleavage sites, black arrowheads mark cleavages possibly performed by a cellular proteinase.

conserved among EAV, LDV, and PRRSV. In contrast, the middle portions differ markedly in size (210–670 residues) and sequence (37–39) (Fig. 3), suggesting that nsP2 has species-specific rather than genus-specific functions (94). Multiple sequence alignments suggest that the nsP2/nsP3 cleavage sites for LDV and PRRSV are Gly–Gly at positions 1286/1287 and 1462/1463, respectively.

Inhibition of cleavage at the nsP2/3 junction abolishes downstream proteolytic events, which are probably all mediated by a 3C-like serine protease (sp) (98) located within nsP4. Site-directed mutagenesis results suggest that the catalytic triad of the nsP4 protease comprises His¹¹⁰³, Asp¹¹²⁹, and Ser¹¹⁸⁴, while Thr¹¹⁷⁹ and His¹¹⁹⁸ may be involved in substrate recognition. Snijder *et al.* (98) further identified three cleavage sites within POL1a (E¹⁰⁶⁴/G¹⁰⁶⁵, E¹²⁶⁸/S¹²⁶⁹, and E¹⁶⁷⁷/G¹⁶⁷⁸) and two additional cleavage sites were proposed in the C-terminus of POL1a (99). The corresponding cleavage sites in LDV and PRRSV in Fig. 3 are inferred.

Three putative recognition sequences for the nsP4 protease were predicted in POL1b. Proteolytic cleavage at these sites would separate the RdRp motif from the putative metal binding and helicase domains. Reaction with specific antisera detected four possible cleavage products designated p80, p50, p26, and p12, respectively (Fig. 3), and a number of putative precursor proteins in lysates of EAV-infected cells (99). The most N-terminal cleavage product, p80, contains the RdRp domain, and the putative zinc finger and helicase motifs are in the adjacent p50. The CVL domain (motif 3; Fig. 1b) is in p26.

Nidovirales Polyprotein Processing: Differences and Common Concepts

No information is available on the processing of POL1b of toroviruses, although the sequence contains a number of potential 3clp- cleavage sites. Because the POL1b sequences of toro- and coronaviruses are colinear (Fig. 1b), the processing of torovirus POL1b is likely to be very similar to that of coronaviruses. There are some marked differences between Coronaviridae and Arteriviridae. The latter lack a cleavage product containing motif 1 (Figs 1b and 1c). Moreover, it remains to be seen whether the C-terminal POL1b cleavage products of the Arteri- and Coronaviridae are functionally equivalent.

For the arteri- and coronaviruses, POL1b processing would yield a product containing both the helicase domain and the zinc finger motif. Such a combination

is rare, but not unprecedented as it has also been seen in *glh-1*, a putative RNA helicase from *Caenorhabditis elegans* (100), and the (putative) yeast RNA helicases Yer176W (101) and NAM7 (102,103). Most helicases lack zinc finger motifs, and it is therefore unlikely that the zinc fingers are required for helicase activity (100). Perhaps, they may confer sequence specificity, for example, in promoter recognition.

GENES EXPRESSED FROM SUBGENOMIC mRNAs

ORFs and Coding Assignments

The arteriviruses PRRSV, LDV, and EAV each possess six genes, numbered 2–7 from the 5' end, that are expressed from subgenomic mRNAs (37–39,44). These ORFs usually overlap (Fig. 1a). ORFs 2, 5, 6, and 7 are conserved among all arteriviruses and, using EAV terminology, code for G_S, G_L, M, and N, respectively (21,22,24,104,105). Sequence similarity can be detected only at the amino acid level; the conservation is generally low and, especially in the EAV proteins, restricted to short domains. ORFs 3 and 4 are conserved among PRRSV, LDV, and SHFV and code for membrane glycoproteins, which in the case of PRRSV, are present in purified virions (106,107). The ORF4 product of EAV shares no obvious sequence similarity with that of the other arteriviruses and has not been detected in virus preparations. Surprisingly, SHFV possesses three additional ORFs. From the limited sequence similarities and the apparent positional conservation of cysteine residues it appears that these ORFs have arisen from a heterologous RNA recombination event by which ORFs 2–4 were duplicated (E. Godeny, personal communication).

Toroviruses apparently express only four genes from subgenomic mRNAs, all of which encode structural proteins. ETV and BoTV are genetically and serologically closely related and share 84% sequence identity in the 3'-most 3 kb of their genomes (145). PoTV is more distant as judged from the sequence of its nucleocapsid protein, which is only 68% identical to those of the other two viruses (Kroneman *et al.*, unpublished). Snijder *et al.* (108) noted the presence of a small ORF completely contained within the N gene of ETV. This ORF, which would encode a hydrophobic polypeptide of approximately 10 kDa, is conserved in BoTV but abrogated by a termination codon in PoTV.

Coronaviruses possess up to nine ORFs that are expressed from sg mRNAs. Of these, the genes for only

the main structural proteins are conserved among the three subgroups (sequence identities of approximately 30%) as is their relative position in the genome (5' S-E-M-N 3'). Apparently, as coronaviruses diverged, subgroup-specific sets of accessory genes were acquired (5,7,109). For instance, the HE gene and ORF2a, which encodes a cytoplasmic nonstructural phosphoprotein of about 30 kDa (16,110,111) (Fig. 1), are only found in group II viruses. Differences in gene composition occur even among viruses of the same subgroup. In CCV and FCoV, ORFs 7a and 7b are at the 3' end of the genome (112,113), but TGEV, which is serologically and genetically very closely related to CCV and FCoV, lacks 7b (67). HCV 229E lacks both ORFs (68).

All accessory genes tested thus far are dispensable for replication *in vitro* and *in vivo* (16,114–119). The functions of the encoded proteins are poorly understood, but at least some may be involved in virus–host interactions and thus contribute to viral fitness. For example, the 7b gene of FCoV codes for a nonstructural 26-kDa secretory glycoprotein (120). FCoV variants that lack ORF7b readily arise in tissue culture, but among naturally occurring FCoV strains, the gene is strictly maintained and its loss correlates with reduced virulence (118).

In contrast to the other Nidovirales, a number of coronaviruses have polycistronic mRNAs which contain up to three ORFs clustered in a single transcription unit. Downstream ORFs are usually translated by leaky scanning but the synthesis of the E proteins of IBV (ORF 3c) and MHV (ORF 5b) may involve internal initiation of translation mediated by a ribosomal landing pad (5,121–123). The N gene of some group II coronaviruses contains a small internal ORF in the +1 reading frame (Fig. 1) that is expressed in infected cells (24,125). It encodes a hitherto unrecognized structural protein that is not essential for virus replication *in vitro* and *in vivo* (119).

RNA Recombination: A Driving Force in Nidovirales Evolution

The variation in coronavirus gene composition is probably the result of heterologous RNA recombination events during which gene modules (126) were obtained either from nonrelated viruses or from the host. The most compelling example is the HE gene, the product of which is 30% identical to the N-terminal subunit of the hemagglutinin-esterase fusion protein (HEF) of influenza C virus (ICV) (16). Heterologous RNA recombination events must also have taken place

during torovirus evolution. A 0.5-kb remnant of an HE gene was found in the ETV genome (20) and an intact, functional HE gene of 1.2 kb is present in the genome of BoTV (Fig. 1; 145). The torovirus HE protein shares 30% sequence identity with both the influenza C virus HEF and the coronavirus HE. In addition, sequences related to ORF2a of group II coronaviruses were found at the 3' end of ETV ORF1a (20) (Fig. 1). The HE and the ORF2a-related sequences found in corona- and toroviruses were probably not inherited from a common ancestor, but acquired through separate heterologous RNA recombination events (6,20) because (i) the genes are in different positions in the two virus genomes (Fig. 1) and (ii) it is highly unlikely that genes retained during the considerable evolutionary divergence between corona- and toroviruses would have been lost from the genomes of coronavirus subgroups I and III.

The differences among the main structural proteins of the Nidovirales could also be explained by heterologous recombination (3). A switch from an arterivirus-like isometric nucleocapsid structure to the extended helical nucleocapsid structures of the Coronaviridae may have been a determining step in the divergence of the Nidovirales (38). Removal of constraints on genome size would have allowed toro- and coronavirus ancestors to acquire large genomes and thus develop the variation in gene composition seen today. A relatively recent replacement of the N gene may subsequently have led to the divergence of the toro- and coronaviruses.

Homologous RNA recombination (128,129) may also be an important force in Nidovirales evolution. High frequency recombination of coronavirus genomes has been observed in tissue culture (130,131), in experimentally infected animals (132) and in embryonated eggs (133). Homologous recombination allows the rapid exchange of beneficial mutations and also serves as a correction mechanism counteracting Muller's ratchet (134). There is evidence that homologous recombination occurs in IBV genomes in the field (135,136,146) and a genetic exchange between CCV and FCoV serotype I strains may have resulted in the emergence of a new FCoV serotype (118,137,138).

CONCLUDING REMARKS AND FUTURE PERSPECTIVES

The nidoviral replicase module has given rise to viruses that utilize similar replication strategies and yet differ markedly in genetic complexity. Common to

the Nidovirales is the use of a nested set of mRNAs. This property, often regarded as "unique," is shared with the phylogenetically unrelated closteroviruses, a genus of filamentous RNA viruses of plants (139,140). Closteroviruses have genomes of up to 20 kb in length, thus approaching the Coronaviridae in genetic complexity. They also resemble the Nidovirales in genome organization and expression, including the use of large polymerase polyproteins, encoded by two overlapping ORFs located at the 5' end of the genome, and down-regulation of RdRp synthesis by ribosomal frameshifting. These recent findings underscore the power of convergent evolution and indicate that similarities in genomic organization and common mechanisms of gene expression and regulation are not reliable taxonomic criteria by themselves. Even the results of comparative sequence analysis should be regarded with caution. Alignments of RdRp domains have been presented to illustrate the evolutionary relationship between the Nidovirales, but the phylogenetic signal in this domain is not sufficient to support a common ancestry of corona- and arteriviruses (141). Here, the toroviruses provide the "missing link" and thus justify a phylogenetic grouping of corona-, toro-, and arteriviruses (141) (P. Zanotto, personal communication).

The analyses of Nidovirales genomes and the studies on polyprotein processing have led to the identification of many viral proteins, some of which are conserved and some of which are genus- or even species-specific. The next formidable task will be to determine the function of each of these products. What is the added value of the nonconserved POL1a-derived cleavage products? Are they antagonists of the intracellular antiviral response or involved in host shut off? What are the functions of the proteins derived from POL1b? Why are proteins containing motifs 1 and 3 lacking in arteriviruses and what are the consequences for replication and transcription? Are replication and transcription distinct processes? Is there a developmental shift from replication to transcription and if so, how is this regulated? What is the function of the various accessory genes of coronaviruses and how do they contribute to viral fitness? Many of these questions may well be solved in the near future. Both in Leiden (147) and in Utrecht (Glaser *et al.*, in preparation), full-length cDNA clones of the EAV genome have been constructed, from which infectious transcripts can be derived. For coronaviruses, no such clones are yet available. However, homologous RNA recombination can be exploited to introduce site-specific mutations into the viral genome using syn-

thetic (DI) RNAs as donor sequence (65,119,142,143). Targeted RNA recombination provides an attractive strategy to characterize the various *ts*-mutants of MHV (65). Undoubtedly, the recent development of methods to study arteri- and coronaviruses by reverse genetics heralds a new era in Nidovirales research.

ACKNOWLEDGMENTS

The authors thank Drs. E. Godeny for sharing unpublished results and Dr. P. Zanotto for advice and stimulating discussions. We are grateful to Katharina Schick and Jolanda Mijnes for their help in the preparation of the manuscript. R. J. de Groot was supported by a fellowship of the Royal Netherlands Academy for Sciences and Arts.

REFERENCES

- Siddell, S. G. (1995) in *The Coronaviridae* (Siddell, S. G., Ed.), pp. 1–10. Plenum, New York.
- Koopmans, M. and Horzinek, M. C. (1994) *Toroviruses of animals and humans: A review. Adv. Virus Res.* **43**, 233–273.
- Snijder, E. J. and Spaan, W. J. M. (1995) in *The Coronaviridae* (Siddell, S. G., Ed.), pp. 239–255. Plenum, New York.
- Plagemann, P. G. W. and Moennig, V. (1992) Lactate dehydrogenase-elevating virus, equine arteritis virus, and simian hemorrhagic fever virus: A new group of positive-strand RNA viruses. *Adv. Virus Res.* **41**, 99–192.
- Luytjes, W. (1995) in *The Coronaviridae* (Siddell, S. G., Ed.), pp. 33–54. Plenum, New York.
- Snijder, E. J. and Horzinek, M. C. (1995) in *The Coronaviridae* (Siddell, S. G., Ed.), pp. 219–238. Plenum, New York.
- Brown, T. D. K. and Brierley, I. (1995) in *The Coronaviridae* (Siddell, S. G., Ed.), pp. 191–217. Plenum, New York.
- Lai, M. M. C. (1990) Coronavirus: Organization, replication and expression of genome. *Annu. Rev. Microbiol.* **44**, 303–333.
- Lai, M. M. C., Liao, C.-L., Lin, Y.-J., and Zhang, X. (1994) Coronavirus: How a large RNA viral genome is replicated and transcribed. *Inf. Agents Dis.* **3**, 98–105.
- Van der Most, R. G., and Spaan, W. J. M. (1995) in *The Coronaviridae* (Siddell, S. G., Ed.), pp. 11–31. Plenum, New York.
- Weiss, M. and Horzinek, M. C. (1987) The proposed family *Toroviridae*: Agents of enteric infections. *Arch. Virol.* **92**, 1–15.
- Horzinek, M. C. (1981) *Non-arthropod-borne togaviruses*, Academic Press, London.
- Macnaughton, M. R., Davies, H. A., and Nermut, M. V. (1978) Ribonucleoprotein-like structures from coronavirus particles. *J. Gen. Virol.* **39**, 545–549.
- Vennema, H., Godeke, G.-J., Rossen, J. W. A., Voorhout, W. F., Horzinek, M. C., Opstelten, D.-J. E., and Rottier, P. J. M. (1996) Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. *EMBO J.* **15**, 2020–2028.
- Bos, E. C. W., Luytjes, W., Van der Meulen, H., Koerten, H. K., and Spaan, W. J. M. (1996) The production of recombinant infectious DI-particles of a murine coronavirus in the absence of helper virus. *Virology* **218**, 52–60.

16. Luytjes, W., Bredenbeek, P. J., Noten, A. F. H., Horzinek, M. C., and Spaan, W. J. M. (1988) Sequence of mouse hepatitis virus A59 mRNA2: Indications for RNA recombination between coronaviruses and influenza C virus. *Virology* **166**, 415–422.
17. Snijder, E. J., Den Boon, J. A., Horzinek, M. C., and Spaan, W. J. M. (1991) Comparison of the genome organization of toro- and coronaviruses: Both divergence from a common ancestor and RNA recombination have played a role in Berne virus evolution. *Virology* **180**, 448–452.
18. Den Boon, J. A., Snijder, E. J., Krijnse, Locker, J., Horzinek, M. C., and Rottier, P. J. M. (1991) Another triple-spanning envelope protein among intracellularly budding RNA viruses: the torovirus E protein. *Virology* **182**, 655–663.
19. De Groot, R. J., Luytjes, W., Horzinek, M. C., Van der Zeijst, B. A. M., Spaan, W. J. M., and Lenstra, J. A. (1987) Evidence for a coiled-coil structure in the spike proteins of coronaviruses. *J. Mol. Biol.* **196**, 963–966.
20. Snijder, E. J., Den Boon, J. A., Spaan, W. J. M., Weiss, M., and Horzinek, M. C. (1990) Primary structure and post-translational processing of the Berne virus peplomer protein. *Virology* **178**, 355–363.
21. De Vries, A. A. F., Chirnside, E. D., Horzinek, M. C., and Rottier, P. J. M. (1992) Structural proteins of equine arteritis virus. *J. Virol.* **66**, 6294–6303.
22. Faaberg, K. S., and Plagemann, P. G. W. (1995) The envelope proteins of lactate dehydrogenase-elevating virus and their membrane topography. *Virology* **212**, 512–525.
23. Godeny, E. K., Zeng, L., Smith, S. L., and Brinton, M. A. (1995) Molecular characterization of the 3' terminus of the simian hemorrhagic fever virus genome. *J. Virol.* **69**, 2679–2683.
24. Mardassi, H., Massie, B., and Dea, S. (1996) Intracellular synthesis, processing, and transport of proteins encoded by ORFs 5 to 7 of porcine reproductive and respiratory syndrome virus. *Virology* **221**, 98–112.
25. De Vries, A. A. F., Post, S. M., Raamsman, M. J. B., Horzinek, M. C., and Rottier, P. J. M. (1995) The two major envelope proteins of equine arteritis virus associate into disulfide-linked heterodimers. *J. Virol.* **69**, 4668–4674.
26. Faaberg, K. S., Even, C., Palmer, G. A., and Plagemann, P. G. W. (1995) Disulfide bonds between two envelope proteins of lactate dehydrogenase-elevating virus are essential for viral infectivity. *J. Virol.* **69**, 613–617.
27. Notkins, A. L., and Scheele, C. (1963) An infectious nucleic acid from the lactic dehydrogenase agent. *Virology* **20**, 640–642.
28. Lomniczi, B. (1977) Biological properties of avian coronavirus RNA. *J. Gen. Virol.* **36**, 531–533.
29. Snijder, E. J., Ederveen, J., Spaan, W. J. M., Weiss, M., and Horzinek, M. C. (1988) Characterization of Berne virus genomic and messenger RNAs. *J. Gen. Virol.* **69**, 2135–2144.
30. Sagripanti, J. L., Zandomeni, R. O., and Weinmann, R. (1986) The cap structure of simian hemorrhagic fever virion RNA. *Virology* **151**, 146–150.
31. Lai, M. M. C., and Stohman, S. A. (1981) Comparative analysis of RNA genomes of mouse hepatitis viruses. *J. Virol.* **38**, 661–670.
32. Bournsnel, M. E. G., Brown, T. D. K., Foulds, I. J., Green, P. F., Tomley, F. M., and Binns, M. M. (1987) Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. *J. Gen. Virol.* **68**, 57–77.
33. Lee, H.-J., Shieh, C.-K., Gorbalenya, A. E., Koonin, E. V., La Monica, N., Tuler, J., Bagdzhadzhyan, A., and Lai, M. M. C. (1991) The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. *Virology* **180**, 567–582.
34. Herold, J., Raabe, T., Schelle-Prinz, B., and Siddell, S. G. (1993) Nucleotide sequence of the human coronavirus 229E RNA polymerase locus. *Virology* **195**, 680–691.
35. Bonilla, P. J., Gorbalenya, A. E., and Weiss, S. R. (1994) Mouse hepatitis virus strain A59 RNA polymerase gene ORF 1a: Heterogeneity among MHV strains. *Virology* **198**, 736–740.
36. Eleouet, J.-F., Rasschaert, D., Lambert, P., Levy, L., Vende, P., and Laude, H. (1995) Complete sequence (20 kilobases) of the polyprotein-encoding gene 1 of transmissible gastroenteritis virus. *Virology* **206**, 817–822.
37. Den Boon, J. A., Snijder, E. J., Chirnside, E. D., de Vries, A. A. F., Horzinek, M. C., and Spaan, W. J. M. (1991) Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily. *J. Virol.* **65**, 2910–2920.
38. Godeny, E. K., Chen, L., Kumar, S. N., Methven, S. L., Koonin, E. V., and Brinton, M. A. (1993) Complete genomic sequence and phylogenetic analysis of the lactate dehydrogenase-elevating virus (LDV). *Virology* **194**, 585–596.
39. Meulenber, J. J. M., Hulst, M. M., De Meijer, E. J., Moonen, P. L. J. M., Den Besten, A., De Kluyver, E. P., Wensvoort, G., and Moormann, R. J. M. (1993) Lelystad virus, the causative agent of porcine epidemic abortion and respiratory syndrome (PEARS), is related to LDV and EAV. *Virology* **192**, 62–72.
40. Snijder, E. J., Horzinek, M. C., and Spaan, W. J. M. (1990) A 3'-coterminally nested set of independently transcribed messenger RNAs is generated during Berne virus replication. *J. Virol.* **64**, 331–338.
41. Snijder, E. J., Den Boon, J. A., Bredenbeek, P. J., Horzinek, M. C., Rijnbrand, R., and Spaan, W. J. M. (1990) The carboxyl-terminal part of the putative Berne virus polymerase is expressed by ribosomal frameshifting and contains sequence motifs which indicate that toro- and coronaviruses are evolutionarily related. *Nucleic Acids Res.* **18**, 4535–4542.
42. Brierley, I. (1995) Ribosomal frameshifting on viral RNAs. *J. Gen. Virol.* **76**, 1885–1892.
43. Spaan, W., Cavanagh, D., and Horzinek, M. C. (1988) Coronaviruses: Structure and genome expression. *J. Gen. Virol.* **69**, 2939–2952.
44. De Vries, A. A. F., Chirnside, E. D., Bredenbeek, P. J., Gravestien, L. A., Horzinek, M. C., and Spaan, W. J. M. (1990) All subgenomic mRNAs of equine arteritis virus contain a common leader sequence. *Nucleic Acids Res.* **18**, 3241–3247.
45. Sethna, P. B., Hung, S.-L., and Brian, D. A. (1989) Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. *Proc. Natl. Acad. Sci. USA* **86**, 5626–5630.
46. Sethna, P. B., Hofmann, M. A., and Brian, D. A. (1991) Minus-strand copies of replicating coronavirus mRNAs contain anti-leaders. *J. Virol.* **65**, 320–325.
47. Sawicki, S. G., and Sawicki, D. L. (1990) Coronavirus transcription: subgenomic mouse hepatitis virus replicative intermediates function in RNA synthesis. *J. Virol.* **64**: 1050–1056.
48. Chen, Z., Faaberg, K. S., and Plagemann, P. G. W. (1994) Detection of negative-stranded subgenomic RNAs but not of free leader in LDV-infected macrophages. *Virus Res.* **34**, 167–177.
49. Den Boon, J. A., Kleijnen, M. F., Spaan, W. J. M., and Snijder, E. J. (1996) Equine arteritis virus subgenomic mRNA synthesis: Analysis of leader-body junctions and replicative-form RNAs. *J. Virol.* **70**, 4291–4298.
50. Hiscox, J. A., Mawditt, K. L., Cavanagh, D., and Britton, P. (1995) Investigation of the control of coronavirus subgenomic

- mRNA transcription by using T7-generated negative-sense RNA transcripts. *J. Virol.* **69**, 6219–6227.
51. Jeong, Y. S., Repass, J. F., Kim, Y.-N., Hwang, S.-M., and Makino, S. (1996) Coronavirus transcription mediated by sequences flanking the transcription consensus sequence. *Virology* **217**, 311–322.
 52. Joo, M., and Makino, S. (1995) The effect of two closely inserted transcription consensus sequences on coronavirus transcription. *J. Virol.* **69**, 272–280.
 53. Van Marle, G., Luytjes, W., Van der Most, R. G., Van der Straaten, T., and Spaan, W. J. M. (1995) Regulation of coronavirus mRNA transcription. *J. Virol.* **69**, 7851–7856.
 54. Krishnan, R., Chang, R.-Y., and Brian, D. A. (1996) Tandem placement of a coronavirus promoter results in enhanced mRNA synthesis from the downstream-most initiation site. *Virology* **218**, 400–405.
 55. Furuya, T., and Lai, M. M. C. (1993) Three different cellular proteins bind to complementary sites on the 5'-end-positive and 3'-end-negative strands of mouse hepatitis virus RNA. *J. Virol.* **67**, 7215–7222.
 56. Zhang, X., and Lai, M. M. C. (1995) Interactions between the cytoplasmic proteins and the intergenic (promoter) sequence of mouse hepatitis virus RNA: Correlation with the amounts of subgenomic mRNA transcribed. *J. Virol.* **69**, 1637–1644.
 57. Chen, Z., Kuo, L., Rowland, R. R., Even, C., Faaborg, K. S., and Plegemann, P. G. W. (1993) Sequences of 3' end of genome and of 5' end of open reading frame 1a of lactate dehydrogenase-elevating virus and common junction motifs between 5' leader and bodies of seven subgenomic mRNAs. *J. Gen. Virol.* **74**, 643–660.
 58. Meulenbergh, J. J. M., De Meijer, E. J., and Moormann, R. J. M. (1993) Subgenomic RNAs of Lelystad virus contain a conserved leader-body junction sequence. *J. Gen. Virol.* **74**, 1697–1701.
 59. Snijder, E. J., Den Boon, J. A., Horzinek, M. C., and Spaan, W. J. M. (1991) Characterization of defective interfering Berne virus RNAs. *J. Gen. Virol.* **72**, 1635–1643.
 60. Van der Most, R. G., De Groot, R. J., and Spaan, W. J. M. (1994) Subgenomic RNA synthesis directed by a synthetic defective interfering RNA of mouse hepatitis virus: A study of coronavirus transcription initiation. *J. Virol.* **65**, 3656–3666.
 61. Tahara, S. M., Dietlin, T. A., Bergmann, C. C., Nelson, G. W., Kyuwa, S., Anthony R. P., and Stohlman, S. A. (1994) Coronavirus translational regulation: Leader affects mRNA efficiency. *Virology* **202**, 621–630.
 62. Luytjes, W., Gerritsma, H., and Spaan, W. J. M. (1996) Replication of synthetic defective interfering RNAs derived from coronavirus mouse hepatitis virus-A59. *Virology* **216**, 174–183.
 63. Kim, Y.-N., Jeong, Y. S., and Makino, S. (1993) Analysis of cis-acting sequences essential for defective interfering RNA replication. *Virology* **197**, 53–63.
 64. Chang, R. Y., Hofmann, M. A., Sethna, P. B., and Brian, D. A. (1994) A cis-acting function for the coronavirus leader in defective interfering RNA replication. *J. Virol.* **68**, 8223–8231.
 65. Masters, P. S., Koetzner, C. A., Kerr, C. A., and Heo, Y. (1994) Optimization of targeted RNA recombination and mapping of a novel nucleocapsid gene mutation in the coronavirus mouse hepatitis virus. *J. Virol.* **68**, 328–337.
 66. Lin, Y.-J., and Lai, M. M. C. (1993) Deletion mapping of a mouse hepatitis virus defective interfering RNA reveals the requirement of an internal and discontinuous sequence for replication. *J. Virol.* **67**, 6110–6118.
 67. Kapke, P. A., and Brian, D. A. (1986) Sequence analysis of the porcine transmissible gastroenteritis coronavirus nucleocapsid protein gene. *Virology* **151**, 41–49.
 68. Schreiber, S. S., Kamahora, T., and Lai, M. M. C. (1989) Sequence analysis of the nucleocapsid protein gene of human coronavirus 229E. *Virology* **169**, 142–151.
 69. Lin, Y.-J., Liao, C.-L., and Lai, M. M. C. (1994) Identification of the cis-acting signal for minus-strand RNA synthesis of a murine coronavirus: Implications for the role of minus-strand RNA in RNA replication and transcription. *J. Virol.* **68**, 8131–8140.
 70. Zeng, L., Godeny, E. K., Methven, S. L., and Brinton, M. A. (1995) Analysis of simian hemorrhagic fever virus (SHFV) subgenomic RNAs, junction sequences, and 5' leader. *Virology* **207**, 543–548.
 71. Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P., and Blinov, V. M. (1989) Coronavirus genome: Prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. *Nucleic Acids Res.* **17**, 4847–4861.
 72. Denison, M. R., and Perlman, S. (1986) Translation and processing of mouse hepatitis virus virion RNA in a cell-free system. *J. Virol.* **60**, 12–18.
 73. Denison, M., and Perlman, S. (1987) Identification of a putative polymerase gene product in cells infected with murine coronavirus A59. *Virology* **157**, 565–568.
 74. Soe, L. H., Shieh, C.-K., Baker, S. C., Chang, M.-F., and Lai, M. M. C. (1987) Sequence and translation of the murine coronavirus 5'-end genomic RNA reveals the N-terminal structure of the putative RNA polymerase. *J. Virol.* **61**, 3968–3976.
 75. Baker, S. C., Shieh, C.-K., Soe, L. H., Chang, M.-F., Vannier, D. M., and Lai, M. M. C. (1989) Identification of a domain required for autoproteolytic cleavage of murine coronavirus gene A polyprotein. *J. Virol.* **63**, 3693–3699.
 76. Bonilla, P. J., Hughes, S. A., Piñón, J. D., and Weiss, S. R. (1995) Characterization of the leader papain-like proteinase of MHV-A59: Identification of a new in vitro cleavage site. *Virology*, **209**, 489–497.
 77. Baker, S. C., Yokomori, K., Dong, S., Carlisle, R., Gorbalenya, A. E., Koonin, E. V., and Lai, M. M. C. (1993) Identification of the catalytic sites of a papain-like cysteine proteinase of murine coronavirus. *J. Virol.* **67**, 6056–6063.
 78. Dong, S., and Baker, S. C. (1994) Determinants of the p28 cleavage site recognized by the first papain-like cysteine proteinase of murine coronavirus. *Virology* **204**, 541–549.
 79. Hughes, S. A., Bonilla, P. J., and Weiss, S. R. (1995) Identification of the murine coronavirus p28 cleavage site. *J. Virol.* **69**, 809–813.
 80. Denison, M. R., Zoltick, P. W., Hughes, S. A., Giangreco, B., Olson, A. L., Perlman, S., Leibowitz, J. L., and Weiss, S. R. (1992) Intracellular processing of the N-terminal ORF 1a proteins of the coronavirus MHV-A59 requires multiple proteolytic events. *Virology* **189**, 274–284.
 81. Denison, M. R., Hughes, S. A., and Weiss, S. R. (1995) Identification and characterization of a 65-kDa protein processed from the gene 1 polyprotein of the murine coronavirus MHV-A59. *Virology* **207**, 316–320.
 82. Gao, H.-Q., Schiller, J. J., and Baker, S. C. (1996) Identification of the polymerase polyprotein products p72 and p65 of the murine coronavirus MHV-JHM. *Virus Res.* **45**, 101–109.
 83. Liu, D. X., Tibbles, K. W., Cavanagh, D., Brown, T. D. K., and Brierley, I. (1995a) Identification, expression, and processing of an 87-kDa polypeptide encoded by ORF 1a of the coronavirus infectious bronchitis virus. *Virology* **208**, 48–57.

84. Gorbalenya, A. E., Donchenko, A. P., Blinov, V. M., and Koonin, E. V. (1989) Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. *FEBS Lett.* **243**, 103–114.
85. Strauss, J. H. Ed. (1990) Viral proteinases. *Semin. Virol.* **1**, 307–384.
86. Tibbles, K. W., Brierley, I., Cavanagh, D., and Brown, T. D. K. (1996) Characterization in vitro of an autocatalytic processing activity associated with the predicted 3C-like proteinase domain of the coronavirus avian infectious bronchitis virus. *J. Virol.* **70**, 1923–1930.
87. Lu, Y., Lu, X., and Denison, M. R. (1995) Identification and characterization of a serine-like proteinase of the murine coronavirus MHV-A59. *J. Virol.* **69**, 3554–3559.
88. Lu, X., Lu, Y., and Denison, M. R. (1996) Intracellular and in vitro-translated 27-kDa proteins contain the 3C-like proteinase activity of the coronavirus MHV-A59. *Virology* **222**, 375–382.
89. Ziebuhr, J., Herold, J., and Siddell, S. G. (1995) Characterization of a human coronavirus (strain 229E) 3C-like proteinase activity. *J. Virol.* **69**, 4331–4338.
90. Liu, D. X., Brierley, I., Tibbles, K. W., and Brown, T. D. K. (1994) A 100-kilodalton polypeptide encoded by open reading frame (ORF) 1b of the coronavirus infectious bronchitis virus is processed by ORF 1a products. *J. Virol.* **68**, 5772–5780.
91. Liu, D. X., and Brown, T. D. K. (1995) Characterisation and mutational analysis of an ORF 1a-encoding proteinase domain responsible for proteolytic processing of the infectious bronchitis virus 1a/1b polyprotein. *Virology* **209**, 420–427.
92. Grötzinger, C., Heusipp, G., Ziebuhr, J., Harms, U., Süß, J., and Siddell, S. G. (1996) Characterization of a 105-kDa polypeptide encoded in gene 1 of the human coronavirus HCV 229E. *Virology* **222**, 227–235.
93. Bredenbeek, P. J., Snijder, E. J., Noten, A. F. H., Den Boon, J. A., Schaaper, W. M. M., Horzinek M. C., and Spaan, W. J. M. (1990) The polymerase gene of corona- and toroviruses: Evidence for an evolutionary relationship. *Adv. Exp. Med. Biol.* **276**, 307–316.
94. Den Boon, J. A., Faaberg, K. S., Meulenberg, J. J. M., Wassenaar, A. L. M., Plagemann, P. G. W., Gorbalenya, A. E., and Snijder, E. J. (1995) Processing and evolution of the N-terminal region of the arterivirus replicase ORF1a protein: Identification of two papainlike cysteine proteases. *J. Virol.* **69**, 4500–4505.
95. Snijder, E. J., Wassenaar, A. L. M., and Spaan, W. J. M. (1992) The 5' end of the equine arteritis virus replicase gene encodes a papainlike cysteine protease. *J. Virol.* **66**, 7040–7048.
96. Snijder, E. J., Wassenaar, A. L. M., and Spaan, W. J. M., (1994) Proteolytic processing of the replicase ORF1a protein of equine arteritis virus. *J. Virol.* **68**, 5755–5764.
97. Snijder, E. J., Wassenaar, A. L. M., Spaan, W. J. M., and Gorbalenya, A. E. (1995) The arterivirus nsp2 protease: An unusual cysteine protease with primary structure similarities to both papain-like and chymotrypsin-like proteases. *J. Biol. Chem.* **270**, 16671–16676.
98. Snijder, E. J., Wassenaar, A. L. M., Van Dinten, L. C., Spaan, W. J. M., and Gorbalenya A. E. (1996) The arterivirus nsp4 protease is the prototype of a novel group of chymotrypsin-like enzymes, the 3C-like serine proteases. *J. Biol. Chem.* **271**, 4864–4871.
99. Van Dinten, L. C., Wassenaar, A. L. M., Gorbalenya, A. E., Spaan, W. J. M., and Snijder, E. J. (1996) Processing of the equine arteritis virus replicase ORF1b protein: Identification of cleavage products containing the putative viral polymerase and helicase domains. *J. Virol.* **70**, 6625–6633.
100. Roussel, D. L., and Bennett, K. L. (1993) glh-1, a germ-line putative RNA helicase from *Caenorhabditis*, has four zinc fingers. *Proc. Natl. Acad. Sci. USA* **90**, 9300–9304.
101. Mulligan, J. T., Dietrich, F. S., Hennessey, K. M., Sehl, P., Komp, C., Wei, Y., Taylor, P., Nakahara, K., Roberts D., and Davis, R. W. (1993) Swiss Protein Database, Accession No. p32644.
102. Leeds, P., Wood, J. M., Lee, B. S., and Culbertson, M. R. (1992) Gene products that promote mRNA turnover in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **12**, 2165–2177.
103. Altamura, N., Groudinsky, O., Dujardin, G., Slonimski, P. P. (1992) NAM7 nuclear gene encodes a novel member of a family of helicases with a Zn-ligand motif and is involved in mitochondrial functions in *Saccharomyces cerevisiae*. *J. Mol. Biol.* **224**, 575–587.
104. Meulenberg J. J. M., Petersen-Den Besten, A., De Kluyver, E. P., Moorman, R. J. M., Schaaper, W. M. M., and Wensvoort, G. (1995) Characterization of proteins encoded by ORFs 2 to 7 of Lelystad virus. *Virology* **206**, 155–163.
105. Meulenberg, J. J. M., and Petersen-Den Besten, A. (1996) Identification and characterization of a sixth structural protein of Lelystad virus: The glycoprotein GP2 encoded by ORF2 is incorporated in virus particles. *Virology* **225**, 44–51.
106. Van Nieuwstadt, A. P., Meulenberg, J. J. M., Van Essen-Zandbergen, A., Petersen-Den Besten, A., Bende, R. J., Moorman, R. J. M., and Wensvoort, G. (1996) Proteins encoded by open reading frames 3 and 4 of the genome of Lelystad virus (*Arteriviridae*) are structural proteins of the virion. *J. Virol.* **70**, 4767–4772.
107. Wiczorek-Krohmer, M., Weiland, F., Conzelmann, K., Kohl, D., Visser, N., Van Woensel, P., Thiel, H.-J., and Weiland, E. (1996) Porcine reproductive and respiratory syndrome virus (PRRSV): Monoclonal antibodies detect common epitopes on two viral proteins of European and U.S. isolates. *Vet. Microbiol.* **51**, 257–266.
108. Snijder, E. J., Den Boon, J. A., Spaan, W. J. M., Verjans, G. M. G. M., and Horzinek, M. C. (1989) Identification and primary structure of the gene encoding the Berne virus nucleocapsid protein. *J. Gen. Virol.* **70**, 3363–3370.
109. De Groot, R. J., Andeweg, A. C., Horzinek, M. C., and Spaan, W. J. M. (1988) Sequence analysis of the 3' end of the feline coronavirus FIPV 79-1146 genome: Comparison with the genome of porcine coronavirus TGEV reveals large insertions. *Virology* **167**, 370–376.
110. Bredenbeek, P. J., Noten, A. F. H., Horzinek, M. C., and Spaan, W. J. M. (1990) Identification and stability of a 30-kDa non-structural protein encoded by mRNA2 of mouse hepatitis virus in infected cells. *Virology* **175**, 303–306.
111. Cox, G. J., Parker, M. D., and Babiuk, L. A. (1991) Bovine coronavirus nonstructural protein ns2 is a phosphoprotein. *Virology* **185**, 509–512.
112. De Groot, R. J., Ter Haar, R. J., Horzinek M. C., and Van der Zeijst, B. A. M. (1987) Intracellular RNAs of the feline infectious peritonitis coronavirus strain 79-1146. *J. Gen. Virol.* **68**, 995–1002.
113. Horsburgh, B. C., Brierley, I., and Brown, T. D. K. (1992) Analysis of a 9.6 kb sequence from the 3' end of canine coronavirus genomic RNA. *J. Gen. Virol.* **73**, 2849–2862.
114. Wesley, R. D., Woods, R. D., and Cheung, A. K. (1990) Genetic basis for the pathogenesis of transmissible gastroenteritis virus. *J. Virol.* **64**, 4761–4766.
115. Schwarz, B., Routledge, E., and Siddell, S. G. (1990) Murine coronavirus nonstructural protein ns2 is not essential for virus replication in transformed cells. *J. Virol.* **64**, 4784–4791.

116. Weiss, S. R., Zoltick, P. W., and Leibowitz, J. L. (1993) The ns4 gene of mouse hepatitis virus (MHV) strain A59 contains two ORFs and thus differs from ns4 of the JHM and S strains. *Arch. Virol.* **129**, 301–309.
117. Yokomori, K., and Lai, M. M. C. (1991) Mouse hepatitis virus S RNA sequence reveals that nonstructural proteins ns4 and ns5a are not essential for murine coronavirus replication. *J. Virol.* **65**, 5605–5608.
118. Herrewegh, A. A. P. M., Vennema, H., Horzinek M. C., Rottier, P. J. M., and De Groot, R. J. (1995) The molecular genetics of feline coronaviruses: Comparative sequence analysis of the ORF7a/7b transcription unit of different biotypes. *Virology* **212**, 622–631.
119. Fischer, F., Peng, D., Hingley, S. T., Weiss, S. R., and Masters, P. S. (1997) The internal open reading frame within the nucleocapsid gene of mouse hepatitis virus encodes a structural protein that is not essential for viral replication. *J. Virol.* **71**, 996–1003.
120. Vennema, H., Rossen, J. W., Wesseling, J., Horzinek, M. C., and Rottier, P. J. (1992) Genomic organization and expression of the 3' end of the canine and feline enteric coronaviruses. *Virology* **191**, 134–140.
121. Liu, D. X., and Inglis, S. C. (1992) Internal entry of ribosomes on a tricistronic mRNA encoded by infectious bronchitis virus. *J. Virol.* **66**, 6143–6154.
122. Le S-Y., Sonenberg, N., and Maizel, J. V. Jr. (1994) Distinct structural elements and internal entry of ribosomes in mRNA3 encoded by infectious bronchitis virus. *Virology* **198**, 405–411.
123. Thiel, V., and Siddell, S. G. (1994) Internal ribosome entry in the coding region of murine hepatitis virus mRNA 5. *J. Gen. Virol.* **75**, 3041–3046.
124. Lapps, W., Hogue, B. G., and Brian, D. A. (1987) Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein genes. *Virology* **157**, 47–57.
125. Senanayake, S. D., Hofmann, M. A., Maki, J. L., and Brian, D. A. (1992) The nucleocapsid protein gene of bovine coronavirus is bicistronic. *J. Virol.* **66**, 5277–5283.
126. Zimmern, D. (1987) in *RNA Genetics* (Holland, J. J., Domingo, E., and Ahlquist, P., Eds.), pp. 211–240, CRC Press, Boca Raton, FL.
128. Lai, M. M. C. (1992) RNA recombination in animal and plant viruses. *Microbiol. Rev.* **56**, 61–79.
129. Jarvis, T. C., and Kirkegaard, K. (1991) The polymerase in its labyrinth: Mechanisms and implications of RNA recombination. *Trends Genet.* **7**, 186–191.
130. Lai, M. M. C., Baric, R. S., Makino, S., Keck, J. G., Egbert, J., Leibowitz, J. L., and Stohman, S. A. (1985) Recombination between nonsegmented RNA genomes of murine coronaviruses. *J. Virol.* **56**, 449–456.
131. Makino, S., Keck, J. G., Stohman, S. A., and Lai, M. M. C. (1986) High-frequency RNA recombination of murine coronaviruses. *J. Virol.* **57**, 729–737.
132. Keck, J. G., Matsushima, G. K., Makino, S., Fleming, J. O., Vannier, D. M., Stohman, S. A., and Lai M. M. C. (1988) In vivo RNA-RNA recombination of coronavirus in mouse brain. *J. Virol.* **62**, 1810–1813.
133. Kottier, S. A., Cavanagh, D., and Britton, P. (1995) Experimental evidence of recombination in coronavirus infectious bronchitis virus. *Virology* **213**, 569–580.
134. Chao, L. (1990) Fitness of RNA virus decreased by Muller's ratchet. *Nature* **348**, 454–455.
135. Wang, L., Junker, D., and Collisson, E. W. (1993) Evidence of natural recombination within the S1 gene of infectious bronchitis virus. *Virology* **192**, 710–716.
136. Jia, W., Karaca, K., Parrish, C. R., and Naqi, S. A. (1995) A novel variant of avian infectious bronchitis virus resulting from recombination among three different strains. *Arch. Virol.* **140**, 259–271.
137. Motokawa, K., Hohdatsu, T., Aizawa, C., Koyama, H., and Hashimoto, H. (1995) Molecular cloning and sequence determination of the peplomer protein gene of feline infectious peritonitis virus type I. *Arch. Virol.* **140**, 469–480.
138. Vennema, H., Poland, A., Floyd Hawkins, K., and Pedersen, N. C. (1995) A comparison of the genomes of FECVs and FIPVs and what they tell us about the relationships between feline coronaviruses and their evolution. *Feline Practice* **23**, 40–44.
139. Dolja, V. V., Karasev, A. V., and Koonin, E. V. (1994) Molecular biology and evolution of closteroviruses: Sophisticated build-up of large RNA genomes. *Annu. Rev. Phytopathol.* **32**, 261–285.
140. Agranovsky, A. A. (1996) Principles of molecular organization, expression, and evolution of closteroviruses: Over the barriers. *Adv. Virus Res.* **47**, 119–158.
141. Zanutto, P. M. de A., Gibbs, M. J., Gould, E. A., and Holmes, E. C. (1996) A reevaluation of the higher taxonomy of viruses based on RNA polymerases. *J. Virol.* **70**, 6083–6096.
142. Koetzner, C. A., Parker, M. M., Ricard, C. S., Sturman, L. S., and Masters, P. S. (1992) Repair and mutagenesis of the genome of a deletion mutant of the coronavirus mouse hepatitis virus by targeted RNA recombination. *J. Virol.* **66**, 1841–1848.
143. Van der Most, R. G., Heijnen, L., Spaan, W. J. M., and De Groot, R. J. (1992) Homologous RNA recombination allows efficient introduction of site-specific mutations into the genome of coronavirus MHV-A59 via synthetic co-replicating RNAs. *Nucleic Acids Res.* **20**, 3375–3381.
144. Bredenbeek, P. J., Pachuk, C. J., Noten, A. F. H., Charité, Luytjes, W., Weiss, S. R., and Spaan, W. J. M. (1990) The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-A59; a highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. *Nucleic Acids Res.* **18**, 1825–1832.
145. Cornelissen, L. A. H. M., Wierda, C. M. H., Van der Meer, F. J., Herrewegh, A. A. P. M., Horzinek, M. C., Egberink, H. F., and de Groot, R. J. (1997) Hemagglutinin-esterase, a novel structural protein of torovirus. *J. Virol.* **71**(7), in press.
146. Kusters, J. G., Jager, E., Niesters, H. G. M., and Van Der Zeijst, B. A. M. (1990) Sequence evidence for RNA recombination in field isolates of avian coronavirus infectious bronchitis virus. *Vaccine* **8**, 605–608.
147. Van Dinten, L. C., Den Boon, J. A., Wassenaar, A. L. M., Spaan, W. J. M., and Snijder, E. J. (1997) An infectious arterivirus cDNA clone: identification of a replicase point mutation that abolishes discontinuous mRNA transcription. *Proc. Natl. Acad. Sci. USA* **94**, 991–996.